BIO392-cnv-freq

Sofia Pfund

2022-09-28

Step 1: Install packages

```
if (!require(devtools)){ install.packages('devtools')
}
```

```
## Loading required package: devtools
```

```
## Loading required package: usethis
```

```
if (!require(pgxRpi)) { devtools::install_github('progenetix/pgxRpi')
}
```

```
## Loading required package: pgxRpi
```

```
library(pgxRpi)
```

Step2: Search esophageal adenocarcinoma NCIt code:

Step3: Access the CNV frequency data from samples with esophageal adenocarcinoma: C4025 (NCIt website)

Need to add 'codematches' parameter so that only the specific match for the code is returned (don't return all of the child nodes as well).

```
freq <- pgxLoader(type = 'frequency', output = 'pgxseg', filters = 'NCIT:C4025', codematches = TRUE)
```

The retreived data is an object contaning two slots `meta` and `data`.

The `meta` slot looks like this:

```
freq$meta
```

```
##         code                      label sample_count
## 1 NCIT:C4025 Esophageal Adenocarcinoma          865
## 2      total                                    865
```

The `data` slot includes two matrices.

```
names(freq$data)
```

```
## [1] "NCIT:C4025" "total"
```

The frequency matrix looks like this

```
head(freq$data$'NCIT:C4025')
```

```
##      filters reference_name   start     end gain_frequency loss_frequency no
## 1 NCIT:C4025              1       0  400000          8.902          6.127  1
```

```
## 2 NCIT:C4025              1  400000 1400000         13.642          5.665  2
## 3 NCIT:C4025              1 1400000 2400000          9.827          6.243  3
## 4 NCIT:C4025              1 2400000 3400000         13.179         10.983  4
## 5 NCIT:C4025              1 3400000 4400000         12.717         10.058  5
## 6 NCIT:C4025              1 4400000 5400000         10.867         10.636  6
```

Dimension of this matrix

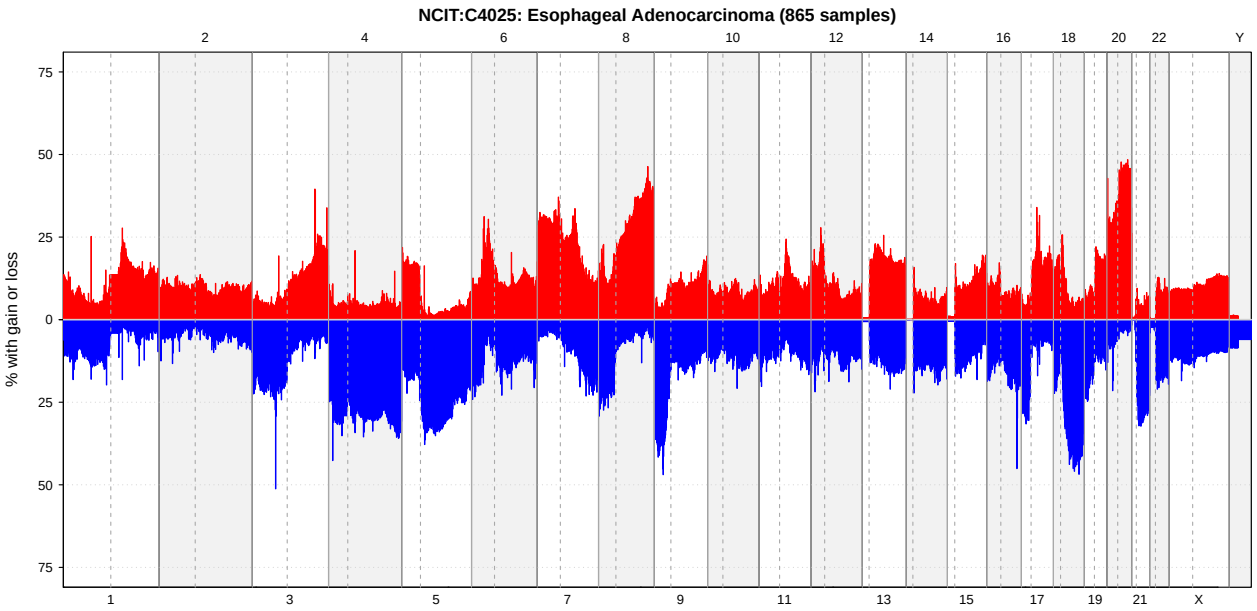
```
dim(freq$data$'NCIT:C4025')
```

```
## [1] 3106    7
```

Step4: Visualize data

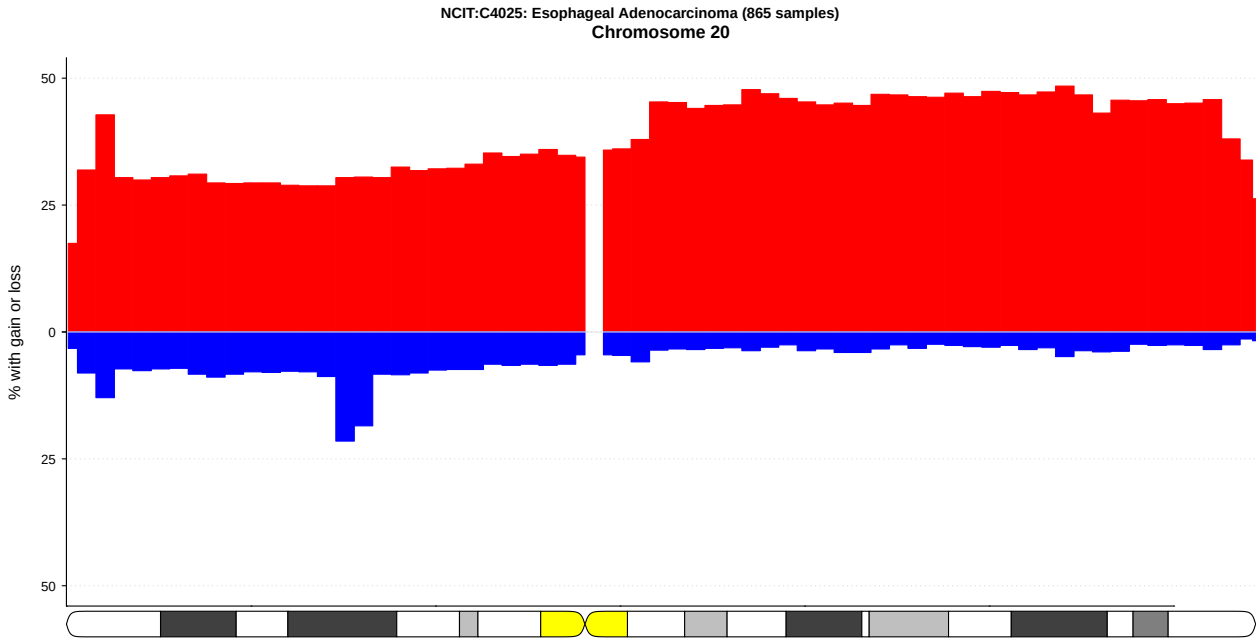
By genome

```
pgxFreqplot(data = freq)
```



By chromosome

```
pgxFreqplot(data = freq, chrom = 20)
```



Step5: Analyse the data

According the plot, we can see frequenct gains on chromosome 7p, 8q, 20p,20q and frequenct losses on chromosome 4p,4q, 5q, 9p, 17p, 18q, 21q.

There is a literature where the findings are consistent with the majority of mine. Here is the paper-link.

A more detailed use case see this link.

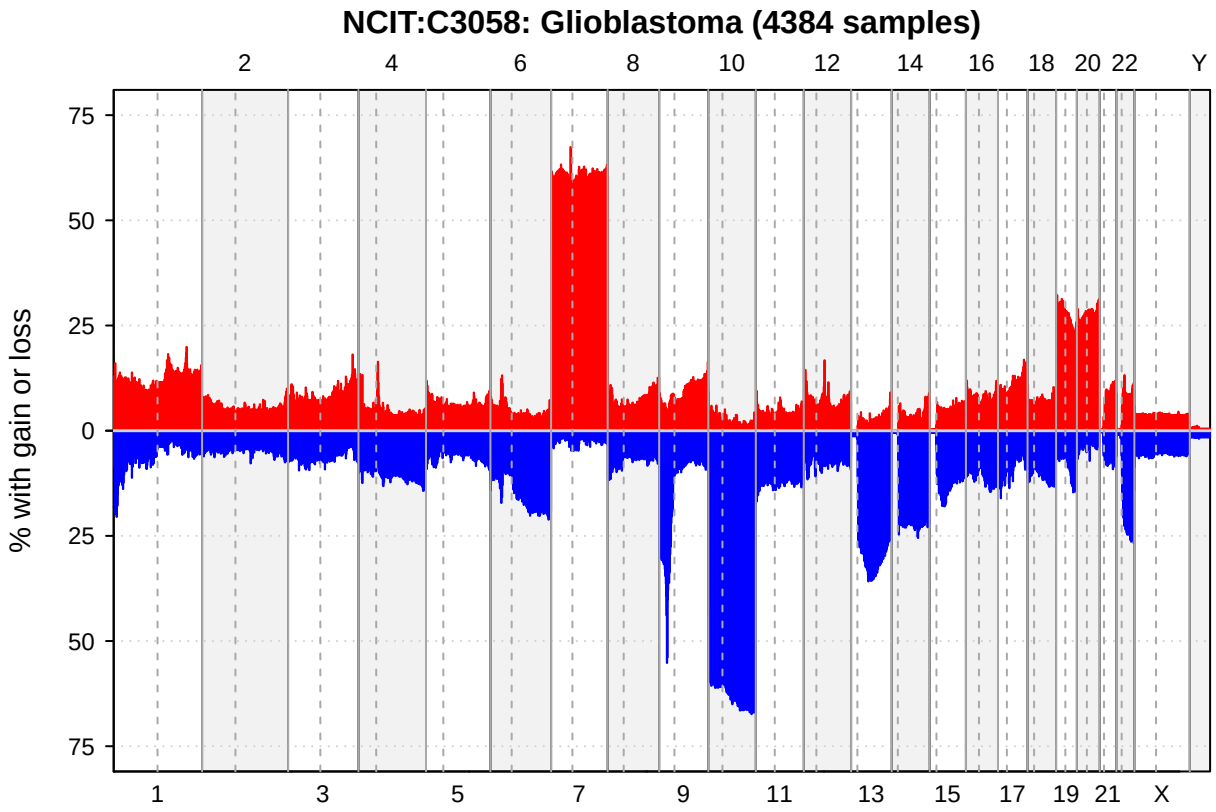
Exercise: check CNV frequecies of other tumor types.

Glioblastoma: NCIT:C3058

```
freq_C3058 <- pgxLoader(type = 'frequency', output = 'pgxseg', filters = 'NCIT:C3058', codematches = TRU
freq_C3058$meta
```

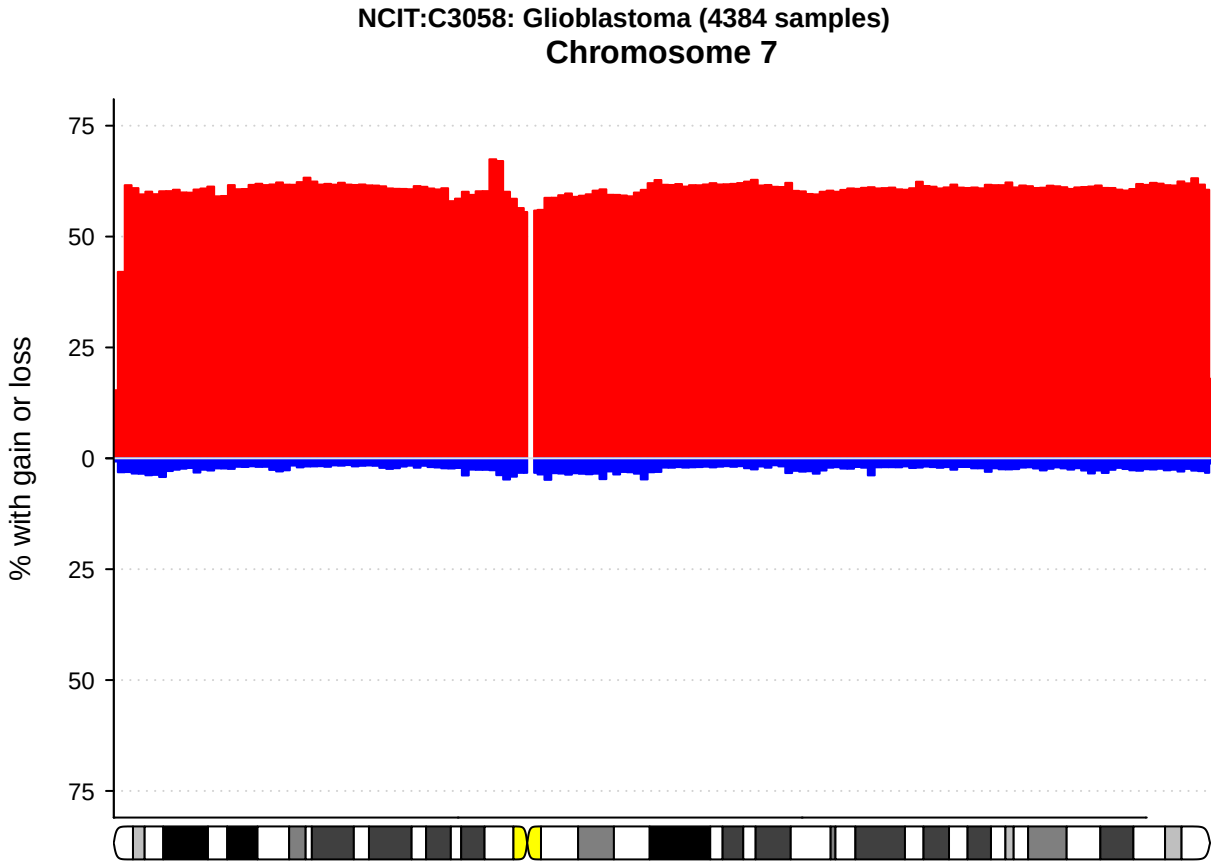
```
##         code        label sample_count
## 1 NCIT:C3058 Glioblastoma         4384
## 2      total                     4384
```

```
pgxFreqplot(data = freq_C3058)
```



In glioblastoma cancer patients there are increased number of copies of chromosome 7, chromosome 19 and chromosome 20 (compared to reference genome). In glioblastoma cancer patients there are fewer copies of chromosome 10 and chromosome 13 (compared to reference genome).

```
pgxFreqplot(data = freq_C3058, chrom = 7)
```


NCIT:C3058: Glioblastoma (4384 samples)
Chromosome 7
% with gain or loss
75
50
25
0
25
50
75